THE RELATION OF REPORTED PREFERENCE TO PERFORMANCE IN PROBLEM SOLVING

by

HERBERT LLOYD BOWMAN, A.B., M.A.

The Abstract of a Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

in the

GRADUATE SCHOOL

of the

UNIVERSITY OF MISSOURI

1929

THE RELATION OF REPORTED PREFERENCE TO PERFORMANCE IN PROBLEM SOLVING

by

HERBERT LLOYD BOWMAN, A.B., M.A.

The Abstract of a Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

in the

GRADUATE SCHOOL

of the

UNIVERSITY OF MISSOURI

1929

*10700

R. UNIV.
BIBLIOTHEEK
LEIDEN

ACKNOWLEDGMENTS

The writer wishes to acknowledge his indebtedness to Professor Theo. W. H. Irion, under whose supervision this study was carried on, for his co-operation, encouragement, and rich counsel; to Dean M. G. Neale and Professor A. G. Capps for suggestions relating to the construction of the research test; to Professor W. J. Saupe for assistance and guidance in the administration of the tests; and to his wife, Ethel Bowman, for assistance in scoring the test forms and in the treatment of the data.

To Mr. O. T. Coil, superintendent of the De Soto Public Schools, to Mrs. J. K. Fyfer, principal of the Jefferson Junior High School of Columbia, and particularly to Superintendent H. U. Hunt of the Sedalia Public Schools, together with the principals and teachers of that system, the writer is indebted for opportunities to carry forward this study, and for the co-operation so graciously given.

H. L. B.

CONTENTS

CHAPTER		PAGE
I.	THE PROBLEM AND ITS BACKGROUND Introduction—The Problem	5
II.	THE RESEARCH INSTRUMENTS EMPLOYED Meaning of Reported Preference—The Research Arithmetic Test—Types of Problems—Reasons for Selection of Types—Arrangement of Problems—Recording Preferences—Selection and Construction of Problems—Other Test Instruments	10
III.	THE SUBJECTS EMPLOYED Grades Employed—Number of Subjects Employed—Intelligence of Subjects—Other Distributions—General Type of Subjects	16
IV.	PROCEDURES Administering the Test—Scoring the Test	20
V.	RESULTS Introduction—Criteria—Measurement of Arithmetic Reading Ability—Teachers' Marks—Research Test Results—Performance—Reported Preference—Preference Scores—Performance and Preference with Respect to Types—Consistency of Preference—Correlations and Comparisons of Test Results—Correlations with Teachers' Marks—Correlations with IQ's	22
VI.	MEANING OF RESULTS Relation of Reported Preference to Performance—Type of Problem Preferred and Intelligence—Intelligence and Performance—Sex Differences in Preference and Performance—Preference—Performance—Grade Differences—Relation of Consistency to Ability—Type of Problem Preferred and Ability—Conclusion	37
VII.	CONCLUSIONS AND SUMMARY Conclusions of Psychological Significance—Practical Conclusions—Further Needed Studies	49
	BIBLIOGRAPHY	52

CHAPTER I

The Problem and Its Background

Introduction

The theory that interest in a learning activity is necessary to pleasurable effort and that through interest success is more probable, is generally accepted among modern educators. Freeland[1] has stated that every educator of importance from Plato and Aristotle to Dewey, Thorndike, and Judd have believed in and practiced some form of the doctrine of interest. A question arises, however, when this theory is assumed to mean that mere expressions or indications of interest in an activity or group of related activities are symptomatic of abilities and consequent achievement in them.

Any expressed interest or preference is considered by some as an unerring guide in the selection of subject matter and learning exercises. The likes and dislikes, it is maintained, are native and permanent in the individual and are indicative of what is best for his well-being and growth. Pringle[2] says,

> "The high place given here to Psychology, and especially Adolescent Psychology, finds support in the doctrines of the "new Education", which teaches that the nature of youth, his active interests and the conditions which surround him should determine each stage of his development, rather than any ideals as to what a mature person should be or do. The question is not what will be of most value to the youth when he grows up, but what is of most interest now."

Continuing, he says,

> "This means that the teacher's chief concern is with the natural adolescent instincts, impulses, intellectual interests, ideals and tendencies of his pupils, rather than with carefully correlated courses of study and logically thought out methods of presentation."

Bobbitt[3] and Kilpatrick[4] also take the view that education should deal with the moving present and that it is living, in the present, with which we should concern ourselves, and consequently that the learner's present interests should be the principal guide in choosing learning activities and curriculum content.

On the other hand, there are educators who take the position that children's interests are varied and changeable, and consequently should not be the sole guide in the selection of subject matter. Counts says, "Nothing should be included in a curriculum merely because it is of interest to children, but whatever is included should be brought into the closest possible relation with

1. Freeland, George E., *Modern Elementary School Practice*, 1925, p. 97.
2. Pringle, Ralph W., *Methods with Adolescents*, 1927, pp. 13-14.
3. Bobbitt, Franklin C., The Orientation of the Curriculum Maker, *Twenty-Sixth Yearbook of the National Society for the Study of Education*, Pt. II, Chap. III, pp. 41-55.
4. Kilpatrick, Wm. H., Statement of Position, *Twenty-Sixth Yearbook of the National Society for the Study of Education*, Pt. II, Chap. X, pp. 119-146.

their interests." He expresses surprise and regret that, "The thought is sometimes suggested that the child should serve as an expert consultant on curriculum content."[5]

A second implication of the question of interests and achievement has to do with methods of teaching. Methods of procedure in the presentation of subject matter are closely tied up with the subject matter itself. The great emphasis placed upon children's interests and preferences as a guide in the choice of learning activities has led consequently to a similar importance being attached to "making things interesting" to the learner. Devices for motivating school work are enumerated in practically all books on teaching technique, and extensively discussed and emphasized in many. The so-called project method is an outstanding example of a procedure in teaching in which efforts are made to utilize children's interests to the greatest possible extent. In support of the project method Kilpatrick[6], who is one of its chief proponents, says that we should not fix in advance just what our pupils should be and think nor the answers they should return to their problems. We should trust the pupils and give them a freer hand.

The scientific basis for the emphasis placed upon the use of children's interests and likes in the choice of subject matter and of methods, is the law of satisfaction, or the law of effect, formulated by Thorndike. This law may be stated in more than one way. A commonly accepted statement of it is, "The ease of learning is increased if either the result or the action used in reaching it is satisfying, but decreased if either the result or the action used in reaching it is annoying."[7] Kilpatrick states this law as one justification of the project method when he says, "A more explicit reason for making the purposeful act the typical unit of instruction is found in the utilization of the laws of learning which this plan affords."[8] Reed further states that there has been no scientific proof established with respect to the validity of the law of effect but that, "we can assume that the law of satisfaction is a good law, and that it is worthwhile to make learning activities as well as end results satisfying and to discover as many devices as possible to make the law effective."[9]

There is a rather distinct division in practices in education with respect to the relative efficacy of the two laws of learning, the law of use and the law of effect. As stated by Symonds and Chase[10] in their recent study of Practice vs. Motivation, there are those who hold that practice and repetition in

5. Counts, George S., Some Notes on the Foundation of Curriculum Making, *Twenty-Sixth Yearbook of the National Society for the Study of Education*, Pt. II, Chap. VI, pp. 73-90.

6. Kilpatrick, Wm. H., *Education for a Changing Civilization*, 1926, pp. 128-129.

7. Reed, Homer B., *Psychology of Elementary School Subjects*, 1927, pp. 23-25.

8. Kilpatrick, Wm. H., *The Project Method*, Teachers College, Columbia University, 1920. pp. 7-9.

9. Reed, Homer B., *loc. cit.*

10. Symonds, Percival M. and Chase, Doris H., Practice vs. Motivation. *Journal of Educational Psychology*, Vol. 20, Jan. 1929, pp. 17-35.

learning are of paramount importance. Motivation is not neglected but it is used only as a device to induce children to practice, for by practice alone is learning accomplished. The other school of thought emphasizes interest and zeal, claiming that by increasing motivation the necessity for repetition may be decreased.

Symonds and Chase experimented with three groups of pupils in the field of learning correct English usage. One group had no motivation outside of the regular school routine. The second group was subjected to test motivation and the third to intrinsic motivation. The conclusion reached following this experiment was that the most important single factor in learning as between repetition and the types of motivation used in this experiment is amount of repetition. Intrinsic motivation such as manufactured by the devices used in this experiment caused no learning in addition to the practice carried on at the same time. Test motivation, in which the pupils were informed of their progress, caused learning over and above that which could be explained by practice. The authors continue, "We must conclude that the most effective device that can be applied to learning is to increase the amount of drill or practice. The prime function of motivation is to make this drill or practice palatable."

Gates[11] in his study on the improvement of reading concluded that there was no evidence that interest was the cause of the pupils' attainments, but on the contrary, that the degree of interest was the result of achievement. Terman[12] reports a correlation of 0.49 between intellectual interest and achievement when the factor of intelligence is held constant. He measured interest by giving arbitrary values to responses made to words used as stimuli.

The question of the relation of interest to ability has been studied principally from the standpoint of broad fields of subject matter. Preferences for certain school subjects or for certain vocational activities have been obtained and compared with achievement in these subjects or fields, the assumption being that verbal expressions of interest are indicative of abilities, and consequently serve as guides for the selection of fields of educational activity, and for vocational guidance. The acknowledgment that interest is symptomatic of ability must be based on the further assumption that interests are not evanescent and highly changeable but are more or less permanent and inherent.

A careful review of these studies reveals that, in most instances, the conception of interest has been that of preference for some one situation over another. It is this meaning of interest we shall employ in this study, using the expression reported preference in order to make the matter still more objective. When an individual prefers one situation to another, we mean

11. Gates, Arthur I., *The Improvement of Reading*, Chap. II, p. 23, 1927.
12. Terman, Lewis M., *Genetic Studies of Genius*, Vol. 1, p. 480. 1925.

that he is more strongly stimulated to respond positively to it than to the second situation. We have no way of measuring the degree of such stimulation but we can draw an inference as to its nature by observing the reactions of the individual. Probably the most straightforward manner of observing is to obtain a report from the individual. In educational practice we frequently rely upon such reports.

THE PROBLEM

In order to throw more light upon the above involved question of interests and the part they play in the learning process, it is believed essential to try to determine the relationship between expressed interest or preference and achievement in a narrow field of activity. This study endeavors to answer this question in part by determining what degree of relationship exists between reported preferences for problems in arithmetic and performance of the problems by pupils of the junior high school level.

Our problem is, then:

1. Are pupils more successful in performing problems for which they express a preference?
2. Of what significance is this relationship of reported preference to performance with respect to:
   a. Nature and types of problems used?
   b. Intelligence and ability of pupils?
   c. Sex?
   d. Grade level of pupils?

In this study, the expression "reported preference" rather than the term "interest" will be employed, since the latter word is more subjective and lends itself to varied interpretations. Reported preference will mean that of several items from which to choose, one item is indicated as being more preferred or better liked than any of the others. It is assumed in this investigation that pupils have preferences as between items of subject matter, that they will indicate these preferences, and that their performances on the items can be measured.

The field of problem solving in arithmetic is one in which such a study seems to be particularly needed and appropriate at this time. Much is being said about the importance of utilizing children's interests as a basis for the construction and selection of descriptive problems for inclusion in textbooks and courses of study. Yet even a casual examination of various arithmetic texts most in use reveals that there is little agreement with respect to types of problems and problem situations included. This fact was brought to attention by Monroe in his report on the construction of arithmetic problem tests.[13]

13. Monroe, Walter S., The Derivation of Reasoning Tests in Arithmetic, *School and Society*, Vol. 8, Sept. 1918, pp. 295-299.

Problem solving in arithmetic can be rather definitely measured, and this fact further influenced the writer to choose this field for investigation.

Some studies which are more or less related to the problem of this investigation have been made during recent years. Conclusions reached from the results of these studies indicate that there is no general agreement as to the relation interest bears to ability, nor as to the stability of earlier interests. Practically all of these investigati ons, however, refer to interests and abilities in broad fields where these traits are general and widely operative rather than specific. This condition favors the entrance of many factors into the measurement of the traits, increasing the subjectivity and consequently the difficulty of accurate measurement and determination. No study in the narrow field of arithmetic problem solving with which this investigation will deal has been reported. Neither have such definite technics been employed to investigate the problem, nor have the measures of preference been as objective and reliable as in this instance.

CHAPTER II

The Research Instruments Employed

As stated in the preceding chapter, the object of this study is to determine and study the relationship between pupils' reported preferences for arithmetic problems and their performance on the problems. This made it necessary to employ tests which would furnish objective measurements of these variables.

MEANING OF REPORTED PREFERENCE

Objective measuring of such factors as interests and attitudes is extremely difficult. Subjectivity enters in to a great extent when we undertake to determine the degree of interest an individual experiences for any situation. For that reason, as stated previously, we shall use the expression reported preference and shall explain it as meaning that, of a group of problems presented to a pupil to be solved, he reports one as being liked better or preferred more than any others of the group.

The question of whether pupils actually have preferences for certain problems or types of problems over others has never been extensively studied. Textbook writers of arithmetic and curriculum makers have proceeded on the assumption that they do. Various types of descriptive problems are included in texts and courses of study in arithmetic for the usually stated purpose of appealing to the various "interests" and "needs" of the pupils. Efforts are made so to clothe the problems in situations that they are made "life-like" and "real". Situations are borrowed from the sciences, from school life, from the home, from business; and these are used as settings for problems.

Our purpose is to determine whether pupils do have and will indicate preferences as among problems, and to learn something of the nature and significance of these indicated preferences.

THE RESEARCH ARITHMETIC TEST

The tests used in this study comprised a research arithmetic test constructed especially for the purpose, and certain standardized tests to be used for analyses and comparisons. In addition to these, the teachers' marks in arithmetic were available as a further criterion. We shall take up a description of the construction of the Research Arithmetic Test at this point.

In undertaking the construction of an adequate research instrument, it was obviously necessary to provide for two things. First, arithmetic problems must be presented to the pupils for actual solution, and second, provision must be made for obtaining reports of preference for certain of these problems over others. Furthermore, these problems of necessity must represent certain definite types throughout the test, in order to provide differentiation and therefore a basis for preference.

There were fifty problems finally selected for use. These were made up into two test forms of twenty-five problems each, which we shall call Form

I and Form II. The two forms were similar in make-up and their results will be combined and treated as one test. Two forms of the test were provided, and given separately, purely for administrative convenience.

Types of Problems

The differentiated types of problems used in the tests were as follows: (1) Problems based upon adult activities; (2) problems based upon children's activities; (3) problems whose setting is in the field of science; (4) problems so stated as to take on the nature of a puzzle; and (5) problems of pure computation only, where the directions for the right procedure are given. For convenience these will be called (1) the adult type, (2) the child type, (3) the science type, (4) the puzzle type, and (5) the computation type.

Reasons for Selection of Types

The Adult Type

Problems based upon adult activities predominated in the older arithmetics, and make up a very large part of the so-called descriptive problems of the texts now in use. The theory supporting this practice is that the child should be introduced to situations in the form of problems which he will meet in later life. This practice still obtains, and textbook makers endeavor to sample as many as possible of the activities engaged in by adults so that the learner will be able to meet them intelligently when he enters upon these activities. The field of adult activity—of the farm or of the store, of the bank or of labor—whatever the case may be, represents a type of problem that is recognized as distinct.

The Child Type

A type of problem situation which is being used increasingly is that problem which has a setting in the every-day activities of the child. As has previously been noted in this study, the view is widely held that children's present interests should be utilized to the fullest extent in the selection of learning exercises. So we find that the child type of problems, together with the adult type mentioned above, make up a great part of the problems used in our arithmetic courses of study.

The Science Type

The science type was chosen not because of its frequent occurrence in arithmetic texts, although this type is found to a certain extent in all of them. It was selected partly for experimental purposes, and partly because it is a type of problem often met in more or less technical literature, in certain occupational activities, and in certain school subjects in science. Furthermore, it is a distinct type in that it involves the use of technical terms and relationships devoid of the human or personal element. Also, if we accept modern theories of transference of training, it will be more necessary to help make transfers to those fields where the subject concerned is used. In

this connection, arithmetic and mathematics are more widely used in the sciences than in any other school subject.

The Puzzle Type

The puzzle problem was included purely for experimental purposes. Few problems of this type are found in modern arithmetic texts. Thorndike believes that they have a place in arithmetic drill, and he includes an exercise of this type in his *Junior High School Mathematics*, Book One, pages 130 to 131. Speaking of the function of this type of problem he says,

> "They have value as drills in analysis of a situation into its elements that wil amuse the gifted children, and as tests of certain abilities. They also require that of many confusing habits, the right one be chosen, rather than that ordinary habits be set aside by some element in the situation."[1]

He believes, furthermore, that if they are given at all they should be presented as puzzles rather than as seeming descriptions of real situations.

The Computation Type

The computation type of problem as used in this study, like the puzzle type, is not a so-called descriptive problem. It differs from the first three types in that no objective material enters into the situation. It is merely put in the form of certain specific directions to be followed, requiring that abstract processes be performed. It differs from the puzzle type in that it contains no puzzle or "catch" element. It also differs from the ordinary drill or pure computation problem found in arithmetics in that operations to be performed are described in words rather than by means of symbols.

ARRANGEMENT OF PROBLEMS

Having described the types of problems used in the test, the arrangement of the problems on the test forms remains to be explained. Each form consisted of six mimeographed pages. Page one contained space for pupil data and instructions to pupils. On the remaining five pages were twenty-five problems to be solved, five problems to a page. One of each of the five types of problems occurred on each page, and the position of the types were rotated so that no type occupied the same location on any two pages.

RECORDING PREFERENCES

At the bottom of each page and beneath the problems, this incompleted sentence was placed to be completed by the pupil: "The problem on this page I like best is No. ____." It was believed that this method of recording the preference increased objectivity, besides providing opportunity for the preference to be made with the situations to be evaluated immediately before the individual, thus increasing the validity of the method.

KEEPING CONSTANT THE FACTOR OF COMPUTATION DIFFICULTY

In order that the preference for a problem be based upon the problem

1. Thorndike, Edward L., *The Psychology of Arithmetic*, 1922, p. 22.

situation and not upon the element of difficulty of computation, it was necessary to hold constant this difficulty factor among the problems of one group. The very nature of this investigation precluded the use of problems that were equal in difficulty with respect to *all* factors. The problem situations, the objects named in the problem, the relative position of the different phrases, the form of the question used, the symbols and the technical terms employed, all affect the difficulty of a problem.[2] These, however, are the very elements which typify the various problems used in this test for comparison. Manifestly a "science" problem could not be classed as such if it contained no terms of a technical nature, nor would a puzzle problem be a puzzle if the phrasing were not such as to put a "catch" in the description. The same elements which affect the relative difficulty of the problems are likewise the elements that in themselves differentiate them into types.

Efforts were made, however, to make the computation difficulty as nearly equal as possible among all the problems of one group or page. This was accomplished in two ways. In the first place, the problems were so selected that all of them on one page required the same operations for correct solution. A second means of equalizing the computation difficulty was that of using as nearly as possible the same digits in the various problems on one page, arranging the digits in different positions, or order. This provided for practically the same combinations of numbers in the steps of the various problems of a group.

Selection and Construction of Problems

In the selection of the problems, it was necessary first that the level of difficulty correspond to the general ability of the pupils to be tested, and second that the problems be representative of the various types to be compared.

It was necessary for the writer to construct most of the problems used in the test. He felt justified in attempting this by reason of several years of experience as a teacher of arithmetic, and through technical knowledge acquired from certain courses related to test construction. Several weeks were spent in a careful examination of a number of modern arithmetic texts and standardized tests. A large list of problems was prepared and these were used to make up two test forms which were given in preliminary try-outs to groups of pupils in two different schools.

A careful analysis of the results of the two try-out tests demonstrated that the difficulty of the problems selected corresponded to the level of ability of average seventh and eighth grade pupils. The uniformity and definiteness with which the preferences were signified, together with the high self correlation of .93 $\pm$ .01 of the tests, encouraged us in the belief that the tests as

2. Hydle, L. L., and Clapp, F. L., Elements of Difficulty in the Interpretation of Concrete Problems in Arithmetic, *University of Wisconsin, Bureau of Education Research Bulletin*, No. 9, Sept. 1927.

finally revised and constructed would prove to be high in objectivity and reliability.

As stated before, the final test used consisted of two test forms of twenty-five problems each, arranged so that there were five problems representing each of the five types, to a page. For each one of the types there were ten chances for reaction. For each page there was given an opportunity for a preference reaction. Thus each subject had the opportunity to make a total of ten preference reactions.

From the results obtained in the final testing, fully reported in Chapter V, we find that when performance scores on Form I are correlated with those of Form II the result is: $r = .90 \pm .005$. Treating all fifty problems as one test and calculating the performance reliability of fifty against fifty reactions $r = .95 \pm .003$. This represents the true performance reliability.

Pursuing now the same procedure with preference scores, the derivation of which is described in Chapter V, we obtain the following: Correlating form against form, $r = .63 \pm .021$; correlating all ten preference reactions against ten reactions, $r = .77 \pm .014$. This will be recognized as a very high r when we take into consideration that we are correlating here only ten reactions with another ten. Such a limited range is always productive of a low r. As will be shown in Chapter V, equalling the number of preference with that of performance items, namely fifty, would give us a coefficient of approximately $.94 \pm .004$. We may, therefore, feel that we have as high a preference score reliability as can be obtained when allowing only ten preference reactions.

Other Test Instruments

As criteria for analysis and comparison, results from certain standardized tests were used. In the fall of 1928, mental tests had been administered to all of the pupils used in this study. Most of them had taken either Form A or Form B of the Otis Self-Administering Test of Mental Ability, Intermediate Examination. A small part of them, however, had been given the Terman Group Test of Mental Ability. These will be treated separately when comparisons with intelligence are made.

Since one trait that was measured in this investigation was performance in arithmetic problems in which the abilities to read and analyze problems are important factors, we felt the need of a standardized test which would serve as a criterion of comparison with respect to these abilities. Stevenson's Arithmetic Reading Test, Form I, Test II, was employed for the purpose.

The pupils of the school where our testing was done had also been examined by Gates' Silent Reading Test, and the results of this test were also available for use in this study. A fourth standardized test whose results were available for our use was the Stanford Achievement Test, Advanced Examination: Form A. Only the results from the two arithmetic tests of this battery will be treated in this study.

The marks in arithmetic given by the teachers at the end of the first semester of the year were obtained to be used as a further criterion of validity. Comparisons and correlations with these measures essential in the analysis of our problem will be found in the appropriate sections of the sequel.

CHAPTER III

The Subjects Employed

The final testing for this investigation was done in the Sedalia Public Schools, Sedalia, Missouri. This was possible through the courtesy of Mr. H. U. Hunt, Superintendent, and the sincere and helpful cooperation of his elementary principals and teachers.

GRADES EMPLOYED

Pupils of the junior high school level were selected as subjects for this study because they represent a stage of educational development that is particularly adapted to such an investigation. It is in this period that the pupils are introduced to so-called finding and exploratory courses on the theory that their interests, their likes and dislikes, are varied and potent and are closely related to their needs and capacities. During the first six years of elementary school training their arithmetic consists uniformly of the acquisition of skills in the fundamental operations, with a certain amount of training in applying those fundamental processes in the solution of problems. Beginning with the seventh grade, however, there is no uniform agreement nor practice with respect to the content of mathematics. In many schools organized on the 6-3-3 plan, general mathematics is given during the seventh, eighth and ninth years, while in others, particularly those of the 8-4 type of organization, the traditional arithmetic is continued. It appears to us that, in view of the efforts being made to formulate subject matter and learning exercises in junior high school mathematics on the basis of pupils' interests and their needs and capacities, a study such as we propose is fitting and urgent at this time.

The Sedalia schools employ the mid-year promotion plan and all available pupils of grades 7B, 7A, 8B, 8A, and 9B were tested. The 7B pupils were just entering upon the work of the 7th grade, and the 9B pupils had just completed the 8th grade when the testing was done. In order to arrive at general conclusions, it is essential that the subjects employed be typical of the grades tested. The following facts are presented to show the typicalness of grades and subjects.

NUMBER OF SUBJECTS EMPLOYED

In the first preliminary test sixty-five pupils were used, thirty-five being in the 8th grade and thirty in the 7th. There were thirty-four 8th grade and twenty-eight 7th grade pupils who were given the second preliminary test.

There was a total of 640 pupils of Sedalia who were given Form I, consisting of 112 seven B, 152 seven A, 98 eight B, 188 eight A, and 90 nine B pupils. The total of 615 were given Form II, divided as follows: 77 seven B, 151 seven A, 101 eight B, 192 eight A, and 94 nine B pupils. The number of

pupils who took both tests were: 72 seven B, 140 seven A, 93 eight B, 172 eight A, and 87 nine B, making a net total of 564 cases available for final calculations. These facts are given in Table 1.

TABLE 1

Number of Pupils Tested and Their Distribution by Half Grades

Grade	Number Taking Test Form I	Number Taking Test Form II	Number Taking Either Form I or Form II, or Both	Number Lost	Number Taking Both Form I and Form II
9B	90	94	97	10	87
8A	188	192	208	36	172
8B	98	101	106	13	93
7A	152	151	163	23	140
7B	112	77	117	45	72
Total	640	615	691	127	564

If we add to the total of 691 who took either Form I or Form II or both, the 128 used in the preliminary testing, we have a grand total of 819 subjects employed in this study.

Intelligence of Subjects

One basis for determining the typical nature of the pupils tested consisted of the distribution of their intelligence expressed in terms of intelligence quotients. Table 2 reveals that the distribution of IQ's of subjects taking both forms of the test was rather normal. The intelligence of 413 of the subjects used in our final calculations was measured by the Otis Self Administering Tests of Mental Ability, Intermediate Examination: Form A, and the intelligence quotients determined according to the directions accompanying this test. A distribution of these makes up Table 2. The Terman Group Test of Mental Ability, Form A, was used in two of the schools of the city and the intelligence quotients of eighty-six pupils used in our final calculations were derived by the method recommended by Terman in this test. The intelligence quotients of twenty-eight of the subjects employed were derived from scores made on the Miller Mental Ability Test, Form A. These will not be used in any calculations where the intelligence of the cases is involved. No intelligence scores could be obtained for the remaining thirty-seven making up the 564 used in our final analyses.

TABLE 2

DISTRIBUTION OF IQ'S ACCORDING TO THE OTIS MENTAL TEST OF 413 PUPILS WHO TOOK FORM I AND FORM II OF THE RESEARCH TEST

IQ	F
145 - 150	2
140 - 145	2
135 - 140	7
130 - 135	15
125 - 130	17
120 - 125	22
115 - 120	36
110 - 115	40
105 - 110	38
100 - 105	38
95 - 100	43
90 - 95	38
85 - 90	32
80 - 85	37
75 - 80	18
70 - 75	14
65 - 70	8
60 - 65	2
55 - 60	4
Total	413
Median	101.38
Mean	101.62
S. D.	17.91
P. E. Mean	0.59

A mean of 101.62 and a median of 101.38 show central tendencies very close to 100, which is an indication of approximate normal intelligence.

OTHER DISTRIBUTIONS

The ages of the pupils at the time Form I was given were obtained and distributions by grades of these ages are typical, the mean age for grade 9B being 14 years, 3.6 months; for grade 8 being 13 years, 10.5 months; and for grade 7 being 12 years, 10.7 months.

The Stanford Achievement Test was given to the pupils of the Sedalia Elementary School during May, 1929. Of the 564 subjects of this study, scores of 529 were obtained. Distributions were made of the scores on the arithmetic tests only and these distributions are quite symmetrical. The mean total score on the two arithmetic tests for 85 Nine B pupils was 229.7 $\pm$ 2.6, for 256 eighth grade pupils, 210.7 $\pm$ 1.7, and for 188 seventh grade pupils, 194.1 $\pm$ 1.7.

General Type of Subjects

Sedalia is a typical midwestern city with a population of 22,745, according to the Bureau of Census estimation of 1925. It has no large factories, its growth and support being due mainly to the surrounding agricultural industry, and to two large railroad shops which furnish employment to several hundred men.

All of the children used in this study in so far as could be determined, were native born and of native born parents. The nationalities represented by the parents are typical of those found in midwestern communities, being principally of English, German, Irish, and Scotch ancestry. The occupational groups are representative, consisting of business, professional, clerical, farming, railroad, and miscellaneous activities.

The foregoing facts and analyses clearly justify the conclusion that the subjects used in this study are typical of the grades they represent.

CHAPTER IV
Procedures

Administering the Test

We have already described in detail the two test forms constructed for our final testing. Each form contained twenty-five problems to be solved, and it was determined in our preliminary try-outs that a period of forty minutes provided adequate time for the completion of one form.

Form I was administered on February 26 and 27, 1929. The pupils tested were distributed over the city in six different schools. The writer administered the test to the majority of pupils. In certain instances where this was not possible, two or three of the teachers of arithmetic gave the test to their respective pupils. The instructions to pupils contained in the test were explicit and clear, as had been shown in the preliminary testing. Care was taken, however, to see that the procedure of administering the test was uniform in all instances.

Form II was identical in mechanical make-up to Form I. The instructions to pupils were the same on both forms as also were the personal data requested. Form II was given on April 16 and 17, 1929. The same schedule with respect to order of school and hour of the day was followed as in the administration of Form I. Likewise, the pupils were tested by the same persons that had given Form I.

The spirit of the pupils who participated in the testing was that of wholesome interest and effort. They were not individuals who had become "test-wise," neither were they unaccustomed to being examined by persons other than their teachers. We believe that the testing was conducted in a highly scientific and objective manner.

Scoring the Arithmetic Test

Each of the two forms was scored for the number of problems correctly solved. This provided for a total possible score of twenty-five on each test, or of fifty on the two combined. The make-up of each page of problems afforded a place for the answer of the problems to be recorded on the right hand margin. This provision made simple the scoring of the papers by means of a key which was constructed for that purpose. No analysis was made of the procedure used by the pupils in solving the problems, and no partial credit for the solutions was given. The problem was merely marked right if the correct answer was obtained, and wrong if the answer was incorrect or if no answer was given.

In recording the preferences from the test forms, cross-section paper was used and the performance and preferences of each pupil were recorded in appropriate columns on the sheet, there being a vertical column for each of the twenty-five problems, as well as one column for the total performance score and

one for the identification number of each pupil. Successful performance of a problem was indicated by the figure 1 in the appropriate square, while a preference was recorded by drawing a small circle in the proper square. Each pupil was credited with a preference for the problem whose number he recorded in the blank provided at the bottom of the page for that purpose. Where no number was indicated, no preference for any problem on that page was scored.

The above remarks refer only to the method of recording the pupils' preferences and performance as taken from the test papers. The method of converting these reported preferences into preference scores will be described in detail in Chapter V.

CHAPTER V
Results

Introduction

In Chapter I of this study the nature of the problem under investigation was discussed together with certain related studies which threw further light upon the question. A description of the research instruments used was given in the second chapter, together with a detailed account of how our research test was devised. Appreciating the importance of basing such a study upon the normality or typicalness of the subjects used, we established the point in Chapter III that the 564 cases to be used in our final calculations and analyses are typical and representative. A description of the testing procedures was made in the fourth and preceding chapter.

We shall now present the results obtained from the various research instruments used. Some comment will be made upon the significance of these results. The full interpretation of them, however, will be reserved for the succeeding chapter.

The results will be taken up under three heads, namely, (1) results obtained from measurements used as criteria for comparisons and analyses, (2) results obtained from the research arithmetic test, and (3) correlations and comparisons between various results of (1) and (2).

Criteria

Results from the Stanford Achievement Test, Advanced Examination: Form A, were obtained, as mentioned in the preceding chapter. Only the combined scores of the two arithmetic tests are included. Distribution of the IQ's are also shown in Chapter IV. (See Table 2). These results will not be repeated at this point, but will be used in making certain comparisons in the third section of this chapter.

Since our Research Test consisted of problems in arithmetic where ability in reading comprehension is one of the factors in performing the problems, we are including scores on Gates' Silent Reading Test, Form A, B, C, and D, as a criterion. This test is designed to measure "general reading ability." In arriving at a single figure to represent an index of average or general reading skill, the "number correct" score for each test was transmuted into a reading grade and the four grades averaged. The distribution of 532 cases is shown in Table 3. Scores were not available for the remaining 32 cases.

As shown in Table 3, the mean reading grade as determined by the Gates Silent Reading Test, was 7.32 for the 532 subjects taking this test. The test was given in the fall of 1928 at the beginning of the school year. The average grade position of the 532 subjects of the test at that time was 7.56. It can be seen that the mean general reading ability as measured by Gates' test is slightly lower than the norm of the test.

TABLE 3

Distribution of 532 Cases in General Reading Ability as Measured by Gates Silent Reading Test for Grades III to VIII

Average Reading Grade	F
11.1 - 12.0	25
10 2 11.1	43
9.3 - 10.2	54
8.4 - 9.3	54
7.5 - 8.4	54
6.6 - 7.5	74
5.7 - 6 6	81
4.8 - 5.7	90
3.9 - 4.8	46
3.0 - 3.9	13
Median	7.04
Mean	7.32
S. D.	2.15
P. E. Mean	0.06

Measurement of Arithmetic Reading Ability

General reading ability, so-called, is made up of many specific abilities. There are, too, probably as many kinds or types of reading as there are purposes in reading, and each of these types has its own group of abilities. The type of reading with which we were especially concerned in this study was that which involves the ability or abilities to read and comprehend an arithmetic problem situation. The Stevenson Arithmetic Reading Test was devised for the purpose of measuring a pupil's ability to comprehend an arithmetic problem and to analyze it into elements. This test was given to 530 of the 564 subjects of our study and the results of this test make up the data contained in Table 4.

The means found for grades 9B, 8 and 7 were 20.01, 18.53 and 16.34, and vary but little from the norms for these grades. The test was given in April, 1929, and the means are as of that date. The test norms of grades 9, 8 and 7 are given as 19.5, 18.7 and 16.4 respectively, and are as of November of the school year.

Teachers' Marks

A generally accepted criterion for validating a test is that of the teachers' marks given on the factor which the test purposes to measure. Distributions by grades of the teachers' marks in arithmetic of 556 pupils make up Table 5. These marks are for the first semester of the school year. The marks, originally expressed in letters, were transformed into comparable numerical quantities in order to make possible their statistical treatment.

TABLE 4

DISTRIBUTION OF 530 CASES ON STEVENSON'S ARITHMETIC READING TEST FORM I, TEST II

	Grade			
	9B	8	7	Total
24 - 26	10	13	2	25
22 - 24	17	44	10	71
20 22	11	49	20	80
18 - 20	24	38	34	96
16 - 18	12	50	41	103
14 - 16	6	28	37	71
12 - 14	3	18	20	41
10 12		8	15	23
8 - 10		5	8	13
6 - 8		2	3	5
4 - 6		1	1	2
Number	83	256	191	530
Median	19.71	18.84	16.56	16.15
Mean	20.01	18.53	16.34	17.97
S. D.	3.19	4.00	3.84	4.05
P. E. Mean	0.24	0.17	0.19	0.12

It can be noticed in Table 5 that the mean teachers' mark for grade 7 is slightly less than that of grade 8, while the latter is not as great as the mean for grade 9B. These differences are probably due in part to the fact that higher grades are usually made up of a smaller percentage of the weaker pupils due to elimination. Ideally, the grade distributions as well as the total distribution of teachers' marks should be normal. The comparatively small number of pupils distributed, however, together with the fact that not all the pupils are marked by the same teacher, would not justify us in expecting a perfectly normal distribution.

RESEARCH TEST RESULTS

We have given only the results obtained from the administration of the various standardized tests used as criteria in this study. Also the teachers' marks in arithmetic were obtained and are shown in Table 5. We shall now present the results obtained from the Research Arithmetic Test devised and used for this investigation.

PERFORMANCE

As has been stated, the two forms of the test constitute, in reality, one test, and will be so treated in this study. There were 564 pupils who took both

TABLE 5

DISTRIBUTION BY GRADES OF 556 CASES ACCORDING TO TEACHERS' MARKS IN ARITHMETIC GIVEN AT THE END OF THE FIRST SEMESTER, 1928-29

Teachers' Marks	Grades			
	9B	8	7	Total
22 - 24		8	6	14
20 - 22	7	11	4	22
18 - 20	6	12	9	27
16 - 18	15	31	22	68
14 - 16	11	25	12	48
12 - 14	14	31	24	69
10 - 12	11	54	52	117
8 - 10	10	32	37	79
6 - 8	6	21	18	45
4 - 6	3	22	14	39
2 - 4	4	10	8	22
0 - 2		1	5	6
Number	87	258	211	556
Median	13.36	11.59	10.90	11.49
Mean	13.14	12.09	11.32	11.96
S. D.	4.54	4.96	4.79	4.90
P. E. Mean	0.33	0.21	0.22	0.14

forms of the test and the score assigned to each pupil will be the sum of the scores on each of the two forms. This provides a possible range of from 0 to 50. The distributions of the total performance on the two forms combined are presented in Table 6.

It can be observed from Table 6 that the scores show a tendency to pile up toward the low score end of the scale. This result was not unexpected. The test was constructed so that its difficulty would prevent many perfect or near perfect scores. Manifestly a test score that is perfect or nearly perfect would yield no basis for comparing performance and preference. Too, a somewhat difficult test in problem solving excludes from operation certain factors, such as the ability to perform a few very simple operations, which would make for the extension of the curve at the lower end. It can be seen also that the range and the variability in performance on the test is somewhat high. This fact can be explained in part by the fact that the subjects of the test ranged from very weak 7B pupils to very strong 9B pupils. A wide range and variability in performance may be expected on a test which attempts to measure subjects where such a range of ability exists. Attention is called to another factor which could not be controlled, due to the nature of the test device used, yet one which

TABLE 6

DISTRIBUTION OF 564 CASES ON BOTH FORMS OF THE RESEARCH ARITHMETIC TEST

Scores	Grades			
	9B	8	7	Total
48 - 52	2	5		7
44 - 48	5	7		12
40 - 44	13	12	4	29
36 - 40	11	33	5	49
32 - 36	11	25	8	44
28 - 32	11	21	11	43
24 - 28	8	22	19	49
20 - 24	9	24	21	54
16 - 20	6	29	27	62
12 16	4	29	31	64
8 12	6	35	40	81
4 - 8	1	19	32	52
0 - 4		4	14	18
Number	87	265	212	564
Median	31.46	22.74	14.58	20.04
Mean	30.28	23.81	16.30	21.99
S. D.	10.93	12.20	9.74	12.18
P. E. Mean	0.79	0.51	0.45	0.35

operated to increase the variability in performance. As has been stated, each of the two forms of our test consisted of twenty-five arithmetic problems. These twenty-five problems were not scaled to represent twenty-five steps or levels of difficulty, but instead were arranged in groups of five problems of approximate equality in difficulty, thus making in reality five main steps of difficulty instead of twenty-five in each test form. Thus, for example, a pupil might perform part or all of the problems of the first group and not have the ability to perform any or many of those of the following. Again, another pupil might succeed very well on two or three groups but might find the next group of problems slightly too difficult. It can be seen, consequently, that a small difference in ability between two pupils might easily be represented by a comparatively wide difference between their scores on each test form and between their total scores on both forms.

Considering the nature of the Research Arithmetic Test and the wide range in arithmetic ability represented by the subjects examined, the scores obtained seem highly satisfactory for our purpose. The reliability of our testing device has already been shown. The validity of the test as a measure of performance in arithmetic problem solving will be discussed in the third section of this chapter.

Reported Preference

The manner in which the subjects indicated their preferences of problems was described in detail in Chapter II, and the method of scoring these preferences is found in the following section of this chapter. A review of the method of obtaining the reports of preferences may not be amiss at this point. As has been stated, the twenty-five problems were arranged in five groups of five problems to a group. The subject was asked to indicate, after working or trying to work the five problems on a page, which one of the five problems he liked best. To do this the pupil merely wrote the number of his preferred problem in a blank at the bottom of the page provided for that purpose. The differentiation between the problems of one group is based upon the objects and the situations or settings which surround or clothe the problems. These settings, as has been explained, are of five kinds or types, namely, (1) the adult type, (2) the child type, (3) the science type, (4) the puzzle type, and (5) the pure computation type. In presenting the results of the record of preferences and in treating these results, we shall deal with individual and total preferences as that of preferences for types of problems, and not as of preferences for particular problems of the fifty included in the test. Our first concern in endeavoring to obtain expressions of preference was the efficacy of our testing device in actually obtaining the preference response. Table 7 shows the result in this connection.

Since there was a total possible preference report per pupil of ten, it can be seen that had there been a perfect report by all pupils, 5640 preferences would have been indicated. The 5222 actually reported represents 92.6% of the total possible.

Due to certain factors, a perfect reporting of preferences was hardly expected. In some cases the pupil would perhaps forget to indicate his preference on certain pages, in others there may have been no actual preference experienced by the subjects, while in still other cases, especially with respect to pupils of low ability, no effort was apparently made to solve some of the groups of problems and as a consequence the preference report was often omitted.

TABLE 7

Total Number of Preferences Possible Together with Total Number Actually Reported by 564 Subjects, by Grades

Grade	Number Pupils	Number Preferences Possible	Number Preferences Reported
9B	87	870	810
8	265	2650	2464
7	212	2120	1948
Tota'	564	5640	5222

Table reads: The 87 pupils of grade 9B had a possible number of 870 preferences, and actually reported 810 preferences, etc.

Altogether the extent to which the pupils reported their preferences was highly satisfactory, not to say surprising. A further expansion and analysis of the facts presented in Table 7 furnish the data for Table 8.

We learn from Table 8 that only twelve pupils of the 564 made but five or fewer reports of preferences, whereas there were 369 who reported the full ten preferences. We are especially interested in those 369 who give us complete preference scores.

TABLE 8

DISTRIBUTION OF THE 564 PUPILS WITH RESPECT TO THE TOTAL NUMBER OF PREFERENCES EACH PUPIL REPORTED

Number Preferences Reported	3	4	5	6	7	8	9	10
Number of Pupils	2	2	8	20	23	63	77	369

Table reads: Out of a possible ten reports of preferences, two pupils made three reports, 2 pupils made four, 8 pupils made five, etc.

PREFERENCE SCORES

Reference was made in Chapter II to the preference scores and the method by which these scores were determined. We shall now describe this method.

In order to obtain an arithmetical value for each preference it was necessary, obviously, to assign to each of the problem types upon which the preferences were based, some arithmetical value. This value or weight was determined by ranking the types according to the total number of preferences reported by all of the subjects for each type of problem. For example, the type for which the greatest number of preferences was reported was ranked first and was assigned the value of one; the type for which the second greatest number of preferences was reported was given the value of two, etc. Table 9 shows the number of times each of the five types of problems was preferred.

TABLE 9

NUMBER OF PROBLEMS PREFERRED OF EACH OF THE FIVE TYPES OF PROBLEMS

	Types of Problems				
	Adult	Child	Science	Puzzle	Computation
No. Problems Preferred	947	1162	799	625	1689

We learn from Table 9 that the computation type ranked first in the number of preferences received, the child type second, the adult type third, the science type fourth, and the puzzle type fifth. Based upon this order of ranking, then, we assigned weights, or values, to each of the types as follows:

computation type.....1
child type.....2
adult type.....3
science type.....4
puzzle type.....5

The next step was actually to determine the preference score for each individual. This was done as follows: The number of preferences for each type of problem was determined for an individual pupil. These numbers were then multiplied, respectively, by the value of the type which they represented. We shall illustrate with an example: Case 13 of grade 8A expressed three preferences for the adult type, five for the child type, and two for the science type. His preference score was determined thus:

3 times 3 = 9
5 times 2 = 10
2 times 4 = 8
Total = 27

In order to insure entire comparability of preference scores, only those of the individuals who expressed ten, or the maximum number of prefer ences were determined. There were 369 pupils who made complete reports of preferences.In making comparisons of preference and performance scores later, only these 369 cases will be used. The preference scores as determined for the 369 cases are shown in distribution in Table 10.

An inspection of Table 10 shows a fairly symmetrical distribution, the median and the mean being 25 80 and 25.41, respectively. A very large part of the scores fall between 18 and 33, there being 302 cases included within this range, with 23 cases above and 44 cases below. Reverting to our description of determining the preference scores, it will be remembered that the computation type of problem had a preference value of one. It can be seen, then, that the lowest possible preference score is ten, for an individual making this type his choice in all ten instances would have a score of ten times one or ten. On the other hand, the maximum score possible would be made when an individual consistently chose the fifth ranking type. In such a situation his score would be fifty, which would represent the highest possible preference score.

The question might be asked as to the meaning or significance of a high or a low preference score. It will be recalled that the weight given to each type was determined by the way the subjects as a group ranked them, the type most preferred being given the value of one, etc. A low preference score would mean, then, that the pupil making it had reported preferences that were in somewhat general agreement with what the group preferred. On the other hand the pupil who scored high in preference necessarily chose in general as his

TABLE 10

DISTRIBUTION OF 369 CASES WITH RESPECT TO PREFERENCE SCORES

Scores	F
40 - 42	3
38 - 40	4
36 - 38	8
34 - 36	8
32 - 34	26
30 32	45
28 - 30	40
26 - 28	46
24 - 26	45
22 - 24	35
20 - 22	30
18 - 20	35
16 - 18	15
14 - 16	11
12 - 14	11
10 - 12	7
Median	25.80
Mean	25.41
S. D.	6.26
P. E. Mean	0.22

preference the types which ranked lower in number of preferences by the group as a whole. Had we ranked the preferences in a reverse order the results would appear differently but would mean precisely what the scores herein reported stand for.

Results of certain correlations of the preference scores with scores on performance will be presented in Chapter VI. It is sufficient at this point to indicate that they are positive and of such magnitude as to lead us to believe that they are not negligible. Evidently our method of scoring has significance and value.

PERFORMANCE AND PREFERENCE WITH RESPECT TO TYPES OF PROBLEMS

The total performance of the 564 pupils on each of the five types of problems was determined and these results, together with the total number of preferences for problems of each type are shown in Table 11.

An inspection of Table 11 seems to indicate that there is considerable variation in the types of problems which pupils like best and in their performance on the problems of the various types. If the pupils had had no distinct preference, and in reporting a preference had merely chosen one at random, then by the laws of chance the number of indicated preferences for each of the five types would have been approximately equal. Likewise, since all of the

TABLE 11

Number of Problems Performed and Number Problems Preferred of Each of the Five Types of Problems

	Types of Problems				
	Adult	Child	Science	Puzzle	Computation
No. Problems Performed	2124	2371	1790	2242	3578
No. Problems Preferred	947	1162	799	625	1689

five problems of one group, representing each of the five types, were of approximately equal computation difficulty, the number of problems of each type correctly solved would have been equal had no factors of difficulty due to the descriptive situations operated. It is evident, therefore, that the situation in which a problem is clothed influences the pupils' preferences and likewise their performances on them. It will be noted that the computation type ranked considerably higher than the others in both preference and performance. It was chosen as a preference more than twice as many times as was the puzzle type, which was the lowest in preference, and it scored over twice as much in performance as the science type, which was the lowest in performance. An analysis of the results pointed to the conclusion that the pupils of low ability in arithmetic rather consistently chose the computation type as a preference. A more complete analysis and further interpretation of the results in this connection will be made in Chapter VI.

Consistency of Preference

We were interested in discovering to what extent the pupils were consistent in choosing a particular type of problem as one most preferred. Do some pupils, for example, distinctly prefer the child type of problem over other types? On the other hand, are there those who clearly have no particular choice as to type? Data in connection with these questions are contained in Table 12 and Table 13. Table 12 groups the pupils according to the greatest number of preferences expressed for any one type. Table 13 groups the pupils with respect to the particular type most preferred by each. Only pupils who reported ten preferences, that is, a preference on each page or group of problems, are included in Table 12.

TABLE 12

Distribution of 369 Pupils Reporting Ten Preferences with Respect to the Greatest Number of Preferences Reported for any One Type

Number Pupils	67	114	82	56	25	11	9	5
Number Preferences	3	4	5	6	7	8	9	10

Table reads: 67 pupils preferred problems of the same type three times, etc.

An inspection of Table 12 indicates that there is some manifestation of a certain degree of consistency of preference, taking the group as a whole. There are 107 pupils who chose the same type of problem six or more times out of the ten. One could hardly arrive at any conclusion regarding the consistency of preference, however, by a mere examination of the data in Table 12. A more extended study by groups, which we shall present in the next chapter, should throw more light upon the question of consistency of preference.

In Table 13 is shown the distribution of 397 pupils with respect to the particular type of problem that was most often preferred. Some of the pupils in this distribution did not report ten preferences, all pupils reporting more preferences for some one type than for any other, being included. There are 167 pupils who seemed to prefer no particular type, reporting the same number of preferences for two or more types. These pupils were not included in this table.

TABLE 13

DISTRIBUTION OF 397 PUPILS WITH RESPECT TO THE TYPE MOST OFTEN PREFERRED

Type	Adult	Child	Science	Puzzle	Computation	Total
Number	43	88	36	26	204	397

Table reads: 43 pupils preferred the adult type of problem most often, etc.

A glance at Table 13 reveals that over one-half of the 397 pupils named the computation type as the one most preferred. Only 26 pupils selected the puzzle problem more often than they did problems of any other type. The child type ranks second with 88 pupils preferring it most, while the adult type ranks third and the science type fourth.

A further analysis and interpretation of the data represented by Table 12 and Table 13 will be made in the next chapter.

CORRELATIONS AND COMPARISONS OF TEST RESULTS

In a study such as this, it is important that the research tests employed be reliable and valid. We have reported in Chapter II a coefficient of reliability of .90 ± .005 between the two forms of this test with respect to performance. When this coefficient was extended by the use of the Spearman-Brown formula, a correlation of .95 ± .003 was obtained for whole test against whole. This indicates a fairly high degree of consistency in measurement of performance. The two forms of the test were likewise correlated with respect to preference scores and a coefficient of .63 ± .02 was obtained. Again applying the Spearman-Brown formula, we found a coefficient of self-correlation of .77 ± .01, which is the coefficient one might expect if the whole test were applied the second time to the same subjects. This is an r of considerable magnitude when we reflect that we are here correlating ten items against ten items, whereas in perform-

ance we are correlating fifty reactions against fifty reactions. Using the Spearman-Brown formula so as to raise the preference reactions to fifty items we get an r of .94. This proves that our preference measures are basically reliable.

As stated a number of times previously, the Research Arithmetic Test was designed to measure two factors, namely, preference for, and performance on arithmetic problems. We have shown in the preceding section that the instrument does obtain reports of preferences. Too, the coefficients of reliability indicate a considerable degree of consistency in measuring the two variables. The question of the validity of the preference scores arises at this point. Do the preferences obtained by our device really represent the pupils' preferences? We endeavored to make clear in the introductory chapter that we were going to be concerned with *reported* preferences. No method has yet been devised that is known to measure objectively an individual's interest or likes as he feels them, if that means anything. There is no measure of subjective factors. We can obtain only the oral or written expression of what the individual *says* that he likes. There is, consequently, no criterion by which the preference scores obtained by our test may be compared for validation. We know only that the individuals reported these preferences, and did so with remarkable consistency. We shall treat those scores merely for what they are, namely, *reported preferences*.

In attempting to establish the validity of our test device with respect to performance in arithmetic problem solving, we have, fortunately, certain criteria which are recognized generally as means of validation. By correlating results obtained from standard measures with those given by the measure to be compared, a coefficient is obtained which serves as an index of validity.

The first comparison we shall present is that of performance on the Research Arithmetic Test with that on the Arithmetic Reasoning Test of the Stanford Achievement test. For the purpose of this comparison we correlated only the scores made on the Stanford Arithmetic Reasoning Test with the scores made on the Research Arithmetic Test. A coefficient of correlation of .82 ± .01 was found between the scores of these two measures. Results from the Research Arithmetic Test were also correlated with those from the Stevenson Arithmetic Reading Test and a coefficient of .72 ± .01 was found between these tests. The results of these calculations are found in Table 14, together

TABLE 14

Correlations of Research Test with Other Tests

	Research Arithmetic Test
Stanford Arithmetic Reasoning Test	.82 ± .01
Stevenson's Arithmetic Reading Test	.72 ± .01
Gates' Silent Reading Test	.46 ± .02

with the correlation discovered between the performance of 532 subjects who took the Gates Silent Reading Test and their performance on the Research Arithmetic Test. The Stevenson test is designed to measure ability to read and analyze arithmetic problems only, there being no calculations called for in the test. Since both the research test and the Stanford test involve computations as well as the ability to read and analyze problems, it was expected that a higher correlation would exist between the Research Arithmetic Test and the Stanford test than between it and the Stevenson test, which proved to be the case. Both correlations, however, seem to indicate a high degree of validity on the part of the research instrument.

The correlation of .46 found between Gates Silent Reading Test and the Research Arithmetic Test indicates that there is a certain relationship between general reading ability as measured by the Gates test, and ability to solve arithmetic problems as measured by the Research Arithmetic Test. This relationship is so low, however, as to indicate that general reading ability is probably made up of many special abilities, and no device designed to measure this general ability in reading can serve to measure any one of the special abilities, one of which is arithmetic problem reading.

Type C of the Gates Silent Reading Test, entitled "Reading to Understand Precise Directions", was constructed for the purpose of measuring ability to read with exactness and precision, such as is necessary in the reading of arithmetic problems. We desired to see what correlation existed between our Research Arithmetic Test and Type C of the Gates test. A coefficient of .41 $\pm$ .102 was obtained when scores from these two measures were correlated. This correlation is not as high as was found between the Research Arithmetic Test and the Gates test as a whole. It is therefore apparent that the Type C test in reading should be further revised and extended before it can be used to measure very accurately the kind of reading involved in solving arithmetic problems, if, indeed, this type of reading does represent the kind most essential in solving problems.

CORRELATION WITH TEACHERS' MARKS

A very commonly used criterion for verifying the validity of a measuring instrument is that of the marks given by the teacher on the trait being measured. We correlated the Research Arithmetic Test results with the teachers' marks in arithmetic given for the first semester of the year in which our testing was done. In order to hold somewhat constant the factor of age and grade, separate correlations were made for each of grades 9B, 8, and 7. The coefficients of correlation with respect to the three separate grades are shown in Table 15. Marks for 556 of the 564 pupils used in our study were available.

It is to be remembered that the scores on the Research Arithmetic Test with which the teachers' marks are correlated represent the measure by this test of ability to solve arithmetic problems, while the marks of the teachers

TABLE 15

Correlations of Research Test with Teachers' Marks in Arithmetic by Grades

Grade	Number Pupils	r
7	211	.52 ± .03
8	258	.62 ± .03
9B	87	.67 ± .04
All Grades	556	.60 ± .02

represent measures of other factors as well. Such factors as attitude and zeal on the part of the pupil, and varying subjective standards of marking among teachers, all operate to influence the evaluation placed upon the pupils' abilities by the teacher. These factors operate to a considerable less degree in the case of an objective test. Consequently a total correlation of .60 between teachers' marks and the Research Arithmetic Test results seems to indicate that certain factors, at least, are being measured by both devices.

Correlation with IQ's

A positive correlation is usually found between the general intelligence of pupils and their achievement in school subjects. Numerous reports and discussions may be found in educational literature in this connection. Since the measurement of arithmetic problem solving makes up only a small part of tests of general intelligence, there is no reason to expect a high correlation between the Research Arithmetic Test used in this study and the Otis and Terman intelligence tests by which the subjects of this study were examined. A positive correlation was expected, and the correlation of .63 ± .02 which involved the 413 pupils who had been examined by the Otis test compares very favorably with results found in comparisons similar to this. The IQ's of eighty-six of the subjects were calculated from results on Terman's test and these were correlated with the arithmetic scores, resulting in a coefficient of correlation of .50 ± .06. The number of cases in this calculation, however, was scarcely enough to afford any great significance to be attached to the resulting coefficient. Table 23 gives the correlation found between the Research Arithmetic Test and the Otis mental test, involving 413 pupils.

TABLE 16

Correlation of Performance on the Research Arithmetic Test with the Otis Mental Test, Involving 413 Cases

	Ot's Test
Research Arithmetic Test	.63 ± .02

A correlation of .63 ± .02 indicates that general intelligence as measured by the Otis test bears some relation to success on the Research Arithmetic Test. Since general reading ability is one of the casual factors determining success on an intelligence test, we desired to discover the relationship existing between arithmetic ability as measured by the Research Arithmetic Test and general intelligence as measured by the Otis test when the factor of general reading ability was held constant. There were 400 pupils who had scores on the Research Arithmetic Test, the Otis mental test and Gates Silent Reading Test. Each of these was correlated with each of the other two. By representing the coefficient of correlation of arithmetic ability and intelligence by r_{12}, that of arithmetic ability and general reading ability by r_{13}, and that of intelligence and general reading ability by r_{23}, the following results were found:

$$r_{12} = .62$$
$$r_{13} = .45$$
$$r_{23} = .61$$

By the use of the partial correlation formula, we obtained the following results:

$$r_{12.3} = .49 \pm .03$$
$$r_{13.2} = .12 \pm .03$$
$$r_{23.1} = .47 \pm .03$$

The above correlations mean that when arithmetic problem solving ability is correlated with intelligence, the factor of general reading ability not operating, the coefficient is .49; when problem solving ability is correlated with general reading ability, intelligence being held constant, the coefficient is .12; and when general reading ability is correlated with intelligence, holding arithmetic ability constant, there is a coefficient of correlation of .47. These coefficients seem to point to the conclusion that there is some relation between general intelligence and problem solving ability, and between general intelligence and general reading ability, as these traits are measured by the respective tests employed. The correlation of .12 found between reading and arithmetic problem solving, however, is too small to be of any predictive value, and the P. E. r too large to make the r reliable. It indicates that among pupils of the same general intelligence, general reading ability as measured by Gates' test is not a causal factor in solving arithmetic problems as measured by the Research Arithmetic Test.

We have presented in this chapter the results obtained from the application of the various test devices used in our study, together with correlations between certain variables. In the next chapter we shall endeavor to extend further the analyses and interpretation of the results contained here.

CHAPTER VI
Meaning of Results

Relation of Reported Preference to Performance

The results obtained in this study were presented in detail in the preceding chapter. We shall now make further interpretations and analyses of them.

The main purpose of our study was to determine what relationship existed between reported preference and performance with respect to arithmetic problem solving by pupils of the junior high school level. We have previously described the method employed in arriving at a preference score. This method, as will be recalled, was based upon the rank given to the five types of problems by the 564 subjects as a whole, and was used because it lends itself to a statistical treatment of the results. A comparison of the preference scores with the performance scores by means of a coefficient of correlation should reveal something of the degree of correspondence between the two variables. There were 369 pupils who furnished scores based upon a complete report of preferences. When these preference scores were correlated with the complete performance scores, a coefficient of correlation of .46 $\pm$ .03 was obtained. Since the preference scores were based upon the rank given by the group to the types of problems, we likewise, and by a similar method, obtained performance scores by weighting the score on each type according to that type's rank in performance by the group. By correlating these two sets of derived scores, a correlation of .58 $\pm$ .02 was discovered.

These correlations suggest that there is some positive relationship between preference and performance as these factors are measured and treated in this study, but the coefficient obtained is manifestly too small to permit reliable prediction. It will be recalled that in determining the preference scores the most frequent preference was given the lowest score value, and the least frequent preference the highest. This method of ranking preference produced a positive coefficient. Should the rank value be reversed by giving the most frequent preference the highest score value, then the coefficient would be negative to the same degree. The significance of the relationship, however, would be the same in either case.

Although, as stated, the degree of relationship does not allow for reliable prediction, yet the coefficient of correlation is markedly positive and considerable significance can be attached to it. It means, in the first place, that there is a positive correspondence between high performance in arithmetic and unusualness of preference, the individuals scoring high on the research test tending to prefer the problems not commonly preferred by the group as a whole. It further indicates that mediocrity in performance goes with mediocrity in preference as valued here, and that low performance corresponds with usualness of preference. In short, as the ability to perform arithmetic problems in-

creases, the preference agrees less and less with the usual preference as determined by the group as a whole.

A majority of the recent investigations, more or less related to this one, reports little or no relationship between preference or interest and performance with respect to rather broad fields of subject matter. We have limited our study to the comparatively narrow activity of arithmetic problem solving. As a consequence the scores which we have correlated are somewhat more specific, being based as they are upon responses to comparatively definite situations. Could still more specific measures of preference and performance responses be obtained in still more definite and specific situations, it is probable that a different correlation between these traits might be found to exist.

The coefficients of correlations reported above were obtained when the preference scores were correlated with total performance. We were interested to learn what correlation existed between preference and performance only on the problems preferred. Accordingly the preference scores were correlated with the actual scores made on the problems that were preferred. We found a coefficient of .22 ± .03 resulting from this correlation. It would seem that one might expect this correlation to be somewhat higher than that between preference and total performance, for in the latter case we were correlating preference with performance not only on preferred problems, but also on those not preferred as well. In using the actual scores only on the preferred problems, however, the range of performance was very narrow, since the maximum number of problems preferred was ten. In this situation it is conceivable that a pupil with a low preference score might easily possess as high a performance score as one with a high preference score. It is probable, then, that the correlation of .22 represents as high or a higher relationship than does the correlation of .46 between preference and total performance.

In order to make more comparable the performance scores on the preferred problems, we weighted each of the preferred problems that was solved and obtained a derived performance score on preferred problems. These were correlated with the preference scores, resulting in a coefficient of correlation of .56 ± .02. This coefficient of correlation is probably the best expression of the relation existing between reported preference and performance. It means that when unusualness of preference is given high value and unusualness of performance is valued in the same way, the correlation between reported preference and performance on the items preferred is as indicated above. Reversing the values both in preference and performance would in this particular calculation produce but little difference.

To summarize these relationships we present the various reputed coefficients in Table 17 on the following page, giving with each coefficient the corresponding Predictive Index as read from Bailor's table.

TABLE 17

A Table Showing all the Coefficients of Correlation Between Reported Preference and Performance Together with the Corresponding Predictive Indices*

	Total Performance Unweighted	Total Performance Weighted	Performance on Preferred Problems Unweighted	Performance on Preferred Problems Weighted
Reported Preference Weighted	.46 ± .03	.58 ± .02	.22 ± .03	.56 ± .02
P. I.	.1121	.1854	.0245	.1716

*The formula for the Predictive Index (P. I.) is P. I. = $1 - \sqrt{1 - r^2}$

Type of Problem Preferred and Intelligence

In the preceding section we based our interpretation upon the group of pupils as a whole. The factor of individual differences, however, is an important one and for that reason we grouped the subjects into certain divisions in order further to analyze the results.

There were 413 subjects whose IQ's had been calculated according to the Otis mental test. These pupils were divided into three groups for the purpose of studying the effect of general intelligence upon preference for types of problems. The pupils were grouped into three divisions, Group 1 consisting of those of the upper quartile, approximately, Group 2 of the middle fifty per cent, and Group 3 of the lower quartile. The scheme following shows the divisions as made.

Group	Range	N
1	All pupils with IQ's of 115 and above	100
2	All pupils with IQ's between 88 and 115	203
3	All pupils with IQ's of 88 and below	110

Table 18 gives the average number of preferences reported by each intelligence group for each of the types of problems.

Attention may be called here to the high S. D.'s reported in this and following tables, as compared with the means of the distributions. These were due to the fact that the distributions of the preferences and performances on one type were by no means normal, being grouped mainly at the lower end of the distribution. Consequently the means are small when compared with the standard deviations. For this reason too much reliance may not be placed in

TABLE 18

MEAN NUMBER OF PREFERENCES REPORTED BY EACH OF THE INTELLIGENCE GROUPS 1, 2, AND 3, ON EACH TYPE OF PROBLEM

Type		Group 1	Group 2	Group 3
Adult	Mean	2.24	2.26	1.92
	S. D.	1.18	1.33	1.06
	P. E. Mean	0.08	0.06	0.07
Child	Mean	2.97	2.48	2.36
	S. D.	1.59	1.53	1.43
	P. E. Mean	0.11	0.07	0.09
Science	Mean	2.41	1.82	1.67
	S. D.	1.46	1.23	1.17
	P. E. Mean	0.10	0.06	0.07
Puzzle	Mean	1.94	1.67	1.30
	S. D	1.48	1.23	1.02
	P. E. Mean	0.10	0.06	0.06
Computation	Mean	2.65	3.62	4.23
	S. D.	2.01	2.26	2.32
	P. E. Mean	0.14	0.11	0.15

the differences of the means. These differences, however, do reveal certain indications or tendencies.

A comparison of the three intelligence groups seems to indicate that those of lower intelligence prefer the descriptive type of problem, represented in the table by the first three types, less than do those of higher general intelligence. Examining the table from left to right, it can be observed that there is a distinct falling off in average number of preferences reported for each of the first four types. On the other hand it can be noted that an increasing number of pupils prefer the computation type as you go from the group of high intelligence to that of low. The computation type of problem as used in this test required no analysis of a problem situation, but consisted merely of directions given as to the calculations to be made. It is possible that this type of problem, in which specific directions were given, has a greater appeal for pupils of low ability. It represents a situation that is more easily understood and which requires no complex analysis for solution. They evidently prefer problems in which they are told in so many words what to do.

If we compare the mean number of preferences which each intelligence group reported for the various types, we learn that there is less variation among those of higher intelligence. Group 1 reported an average of 2.97 preferences for the child type and 1.94 for the puzzle type, a range of 1.03 between

its greatest and least preferred types. On the other hand there is a range of 1.95 for Group 2 and one of 2.95 for Group 3, each of the two groups indicating a greater preference for the computation type and the least for the puzzle type. This difference in variation indicates that the pupils of higher intelligence, due possibly to their greater ability to comprehend the problems, are less concerned about the problem setting than are those of lower intelligence. The preference of intelligent children appears to be based more on the nature of the content of the problem, that is, upon the actual problem situation.

Intelligence and Performance

A comparison of performance on the problem types was made with respect to general intelligence, using the same intelligence groups as were compared in Table 18. The mean performances by the three intelligence groups on each type are presented in Table 19.

We see in Table 19 that there is a uniform decrease in the mean performance of all types as we go from Group I to Group 3. This of course was expected. It is particularly noticeable, however, that the decrease in performance on the puzzle and the science type is considerably greater than that of the other types. For example, the mean performance of Group 1 on the puzzle type is more than four times that of Group 3 on the same type, while it is less than twice as

TABLE 19

Mean Performance by the Intelligence Groups 1, 2, and 3 on Each of the Five Types of Problems

Type		Group 1	Group 2	Group 3
Adult	Mean	6.47	4.26	2.67
	S. D.	2.46	2.43	2.01
	P. E. Mean	0.17	0.12	0.14
Child	Mean	7.18	4.53	2.88
	S. D.	2.41	2.53	2.19
	P. E. Mean	0.16	0.12	0.15
Science	Mean	5.91	3.68	1.87
	S. D.	2.61	2.50	1.80
	P. E. Mean	0.18	0.12	0.12
Puzzle	Mean	7.41	4.59	2.15
	S. D.	2.35	2.94	2.14
	P. E. Mean	0.16	0.14	0.14
Computation	Mean	8.64	7.03	5.15
	S. D.	1.73	2.18	2.30
	P. E. Mean	0.12	0.10	0.15

great on the computation type. Table 20, which gives the per cent of performance on each type, presents a clearer picture of the relative performances.

We learn from Table 20 that the pupils of Group 1, who are of high intelligence, are more uniform in their performance with respect to the different types, while the lowest intelligence group shows the greatest degree of variation. The computation type represents 35.0% of the total number of problems solved by Group 3, while only 12.7% of the number solved by this group are of the science type.

TABLE 20

PER CENT OF TOTAL PERFORMANCE REPRESENTED BY EACH TYPE BY INTELLIGENCE GROUPS

Type	Group 1	Group 2	Group 3
Adult	18.2	17.7	18.1
Child	20.2	18.8	19.6
Science	16.6	15.3	12.7
Puzzle	20.8	19.0	14.6
Computation	24.6	29.2	35.0

Table reads: 18.2% of the problems solved by Group 1 were of the Adult type, 17.7% of those solved by Group 2 were of the Adult type, etc.

A comparison of the rankings of the types of problems in preference and performance by the three intelligence groups reveals to a further extent the variation existing with respect to the two factors. Table 21 gives the order in which the types were preferred and performed, respectively, by the three groups.

It can be observed that the computation type and the child type rank high in preference and performance. The adult type takes a somewhat

TABLE 21

RANK OF TYPES WITH RESPECT TO PREFERENCE AND PERFORMANCE BY EACH OF THE THREE INTELLIGENCE GROUPS

Rank	Group 1		Group 2		Group 3	
	Preference	Performance	Preference	Performance	Preference	Performance
1	child	comp.	comp.	comp.	comp.	comp.
2	comp.	puzzle	child	puzzle	child	child
3	science	child	adult	child	adult	adult
4	adult	adult	science	adult	science	puzzle
5	puzzle	science	puzzle	science	puzzle	science

uniformly medium position, seeming to be slightly more preferred and better performed, relatively, by the lower intelligence groups. On the other hand, performance is distinctly higher on the puzzle type by the groups of higher intelligence, although it ranks last in preference according to all three groups. The science type of problem seems to be somewhat more preferred by the pupils of high intelligence than by those of low, this type ranking lowest in performance, however, with respect to each of the intelligence groups.

Summarizing, an analysis of the results with respect to general intelligence groups brings out the following facts:

1. There is less variation in both preference and performance by the pupils of higher intelligence with respect to types than by those of lower intelligence.
2. Pupils of higher intelligence show a less relative preference for and performance on problems of the computation type than do the pupils of lower intelligence.
3. The science type and puzzle type are distinctly less preferred and performed by the pupils of lower intelligence than are the other three types.
4. There is an increasing positive relation between preference and performance as you go from the group of high intelligence to that of the low.

The above facts seem further to justify the conclusion that the selection of arithmetic problems for use in the junior high school should not be done for the group as a whole but should be done to a certain extent on the basis of intelligence groups.

Sex Differences in Preference and Performance

The question might arise as to whether the girls and boys differ to any extent with respect to what types they prefer and in their performance on them. In this connection we calculated the mean number of preferences reported on each type by each of the two sexes, and also the mean performance on each type.

Preference

There were 315 girls and 249 boys who made up the 564 subjects of our study. The mean number of preferences for each type was calculated for each sex, together with the S. D. and the P. E. Mean. These data are presented in Table 22.

It can be observed that the mean number of preferences reported by the girls for all the types is somewhat in excess of that reported by the boys, the means being 9.90 and 9.57, respectively.

The child type of problem seems to be about equally well liked by both boys and girls. On the other types, however, there appear to be distinct differences as to preference. The boys reported preferences for the adult type and the science type to a greater degree than did the girls, whereas the girls

TABLE 22

MEAN NUMBER OF PREFERENCES REPORTED FOR EACH TYPE OF PROBLEM BY 315 GIRLS AND 249 BOYS

		TYPE					
		Adult	Child	Science	Puzzle	Computation	All Types
Girls	Mean	2.06	2.53	1.72	1.82	3.76	9.90
	S. D.	1.22	1.56	1.23	1.37	2.33	1.13
	P. E. m.	0.05	0.06	0.05	0.05	0.09	0.04
Boys	Mean	2.33	2.60	2.15	1.34	3.16	9.57
	S. D.	1.35	1.43	1.40	0.98	2.44	1.44
	P. E. m.	0.06	0.06	0.06	0.04	0.10	0.06

indicated a greater preference for the puzzle and the computation types than did the boys.

PERFORMANCE

The mean score in performance on each type of problem was calculated for the girls and for the boys, and these data, together with the mean total performance, make up Table 23.

TABLE 23

MEAN NUMBER OF PROBLEMS OF EACH TYPE PERFORMED BY 315 GIRLS AND 249 BOYS

		TYPE					
		Adult	Child	Science	Puzzle	Computation	All Types
Girls	Mean	3.99	4.50	3.37	4.53	7.01	21.39
	S. D.	2.56	2.75	2.58	3.09	2.38	12.02
	P. E. m.	0.10	0.10	0.10	0.12	0.09	0.46
Boys	Mean	4.61	4.96	4.06	4.41	6.64	22.68
	S. D.	2.55	2.75	2.79	3.02	2.42	12.36
	P. E. m.	0.11	0.12	0.12	0.13	0.10	0.53

The mean performance of the boys is slightly higher than that of the girls, a result which had been noted before by other investigators. The P. E. of the difference of the total mean performances was calculated and found to be .702, indicating eighty-nine chances in a hundred that the difference obtained is a true difference.

We learn also from Table 23 that there is a difference between the relative performance of the girls and of the boys upon the various types. It can be

seen that the boys have a higher mean performance on the adult, child and science types, while the girls score higher on the puzzle and computation types. The greatest difference is that of the science type, the boys averaging 3.57 on this type to a mean of 2.87 for the girls.

By combining the data contained in Table 22 and Table 23, we are able better to see the comparison between the sexes with respect to their preferences in relation to performance. Table 24 serves this purpose.

TABLE 24

AVERAGE NUMBER OF PREFERENCES AND AVERAGE NUMBER OF PERFORMANCES BY 315 GIRLS AND 249 BOYS WITH RESPECT TO TYPES OF PROBLEMS

TYPE	GIRLS		BOYS	
	Mean Preference	Mean Performance	Mean Preference	Mean Performance
Adult	2.06	3.99	2.33	4.61
Child	2.53	4.50	2.60	4.96
Science	1.72	3.37	2.15	4.06
Puzzle	1.82	4.33	1.34	4.41
Computation	3.76	7.01	3.16	6.64
Total	9.90	21.39	9.57	22.68

A further analysis of the data making up Table 24 was made for the purpose of observing the order in which the two sexes ranked the types with respect to preference and performance. Table 25 pictures the rankings in this respect.

We learn from Table 24 and Table 25 that the computation type of problem easily ranks first in both preference and performance. This is true for both boys and girls, the latter ranking it higher, however, than the boys.

TABLE 25

RANK OF TYPES OF PROBLEMS WITH RESPECT TO PREFERENCE AND PERFORMANCES DETERMINED BY SCORES MADE BY 315 GIRLS AND 249 BOYS

Rank	TYPE					
	GIRLS		BOYS		BOTH SEXES	
	Preference	Performance	Preference	Performance	Preference	Performance
1	comp.	comp.	comp.	comp.	comp.	comp.
2	child	puzzle	child	child	child	child
3	adult	child	adult	adult	adult	puzzle
4	puzzle	adult	science	puzzle	science	adult
5	science	science	puzzle	science	puzzle	science

The child type of problem is second in point of preference by both boys and girls, the puzzle type, however, having a higher rank in performance by the girls. The adult type is third in preference and fourth in performance by the girls, whereas the boys placed it fourth in preference and third in performance. The girls ranked the puzzle problem fourth in preference, but their performance on this type places it second in rank. The boys showed comparatively less preference for problems of the puzzle type, ranking it lowest in preference. The girls seemed not to prefer the science type, nor did they perform problems of this type as well. The science problems were distinctly more preferred by the boys, and were somewhat better performed, than they were by the girls.

The above analysis and presentation of results indicate that the boys and girls used in this study differ somewhat in the types of problems they prefer and in their performances on them. A rather surprising result was the difference in the mean total performance, the boys exceeding the girls by a 1.29 difference. The inference might be drawn that the settings of the problems represented situations which were more familiar to the boys, and consequently more readily analyzed. This would not explain, however, the girls' apparent superiority in solving the puzzle and pure computation types of problems. A further possible explanation of the superiority of the boys over the girls in performance on the test is that the former represent a somewhat more select group due to a greater elimination of those of lower ability during the junior high school period. The superiority of boys in problem solving, however, has been noted before. In reviewing a number of studies on sex differences in arithmetic, Lincoln[1] found evidence that the boys exceeded the girls in problem solving. He says:

> "In general, it may be said on the basis of the data presented above that girls are somewhat superior to boys in the fundamental operations of arithmetic. When problems are considered the case is not so clear. There is a practically universal agreement that when compared on a grade basis the boys show a greater problem-solving or reasoning ability than the girls."

Further verification of the above was found when we calculated the mean performance of the boys and of the girls on the Arithmetic Reasoning Test of the Stanford Achievement Test. The mean performance of the boys was found to be 87.86 ± .91, while that of the girls was 81.87 with a probable error of .80. There were 251 boys and 308 girls in this calculation.

GRADE DIFFERENCES

In a manner similar to that employed in studying the intelligence groups and the sexes, the subjects were grouped by grades for comparison in preference and performance. There were 87 pupils in grade 9B, 265 in grade 8, and 212 in

1. Lincoln, Edward A., *Sex Differences in the Growth of American Children*, 1927, pp. 56-67.

grade 7. As was the case in the analysis of the intelligence groups, there was an increasing average number of preferences reported for the computation type in going from the higher to the lower grades, while there was a noticeable relative decrease in reported preferences for the science type and the child type. The 9B group ranked the child type as high as the computation type, each averaging 2.97. The range between the lowest preferred type and the highest, as ranked by grade 9B, was 1.27, whereas it was 1.68 for grade 8, and 2.40 for grade 7. This indicates a greater variation or scattering of preferences among the pupils of the higher grades.

There was a consistent decrease in performance on all the types, going from the higher to the lower grades. The difference between the mean performances of grade 9B and grade 7 on each of the types was approximately 2.7, with the exception of the puzzle type. On this type the ninth grade exceeded the seventh by 3.22. This result bears out what we learned in the analysis of the results by intelligence groups, the indication being that pupils of higher ability performed relatively better on problems of this type than those of lower ability.

Relation of Consistency to Ability

In order further to study the relation of consistency in preference to performance, the pupils who reported ten preferences were divided arbitrarily into three groups. The first group included those pupils who reported not more than four preferences for any one type of problem. Since there was none who reported less than three for any one type, Group 1 consisted of all those whose highest preference for one type was three or four. Group 2 consisted of those pupils reporting ten preferences, either five or six of which were for one type of problem. Group 3 included those pupils reporting ten preferences of which either seven, eight, nine, or ten were for one type of problem.

A comparison of these means shows a tendency for both the mean IQ and the mean performance on the Research Arithmetic Test to decrease in going from Group 1 to Group 3. The indication seems to be that those pupils who are more consistent in choosing the same type of problem throughout are lower in general intelligence and in performance on the problems. Or, stating it conversely, the results seem to indicate that pupils of lower ability, as measured here, are more consistent in preferring a particular type of problem, whereas those of higher ability are more scattering and are less concerned as to the type of problem preferred. This is true, however, mainly because of the striking preference shown for the computation problems. Had this type of problem not been included in the test, it is probable that consistency in preference would not have been so distinct.

Type of Problem Preferred and Ability

In the first section of this chapter an analysis of the intelligence groups with respect to preference and performance was presented. Reverting to

Table 19 and Table 20, it will be recalled that these groups differed with respect to the degree of preference manifested for the various types of problems, and also with respect to performance upon them. In order to study this point further, we distributed the pupils according to the particular type of problem they most often preferred. This distribution is shown in Table 13. An inspection of that table shows that 26 pupils showed a greatest preference for the puzzle type of problem, 36 seemed to prefer the science type most, 43 the adult type, 88 the child type, and 204 the computation type. The mean IQ and the mean performance score on the Research Arithmetic Test were calculated for each of these five groups. Although there are not enough cases in two or three of the groups compared to make the means of the IQ's and of the arithmetic performance altogether reliable, we can observe a tendency for the pupils who are low in general intelligence and in performance on the Research Arithmetic Test to prefer the computation type of problem, while those most preferring the science and the puzzle type show the highest mean intelligence and performance scores.

Conclusion

Taking into consideration the limitations of this study, we may note certain tendencies with respect to preference and performance. The coefficients of correlation found between reported preferences and performance in arithmetic problems as presented in the first section of this chapter indicate a somewhat positive relationship between these factors with respect to the group as a whole. This relationship, however, is hardly great enough to be of particular value as a means of prediction.

In making analyses by groups of pupils we find certain definite tendencies. The most significant result shown in this respect is the tendency for pupils of high ability, as measured in this study, to perform equally well on all types of problems, and furthermore to indicate no decided preference for any particular types. On the other hand, pupils of lower ability more consistently prefer the pure computation type and exhibit a higher relative degree of performance on this type of problem. The computation problem used in this test was one which called for simple directions to be read and understood, calling for certain arithmetical calculations to be made. We may draw the conclusion, then, that pupils of lower ability prefer problems which are not set in a complex situation, but which necessitate only the following of definite directions.

The inference may be made, then, that there is a tendency for the individual to select as his preference that problem, out of a group of problems, which he knows or thinks he will be more nearly successful in solving. Within the limitations of this study, we seem to be justified in drawing the further inference that the expectation of greater success leads the individual to prefer one problem over others, and that belief in success causes preference rather than that preference is a cause of successful performance.

CHAPTER VII

Conclusions and Summary

In a study of this type the conclusions naturally fall into certain groups. First, there are those which are of interest primarily to the psychologist. We shall list these as conclusions of psychological significance. A study such as this should also furnish in its conclusions suggestions that would apply specifically to that field in which the investigation was made. In this case such conclusions would relate themselves to arithmetic, and more especially to problem solving in the junior high school. These will be listed as practical conclusions. Finally, there are those conclusions which reveal the necessity for additional researches. These will be listed as further needed studies.

Conclusions of Psychological Significance

Conclusions of psychological significance are of utmost importance in this study. They reveal the true character of certain factors which have been investigated, and explain the relationships which exist between these. Such conclusions frequently seem trivial to the reader in general, but one trained in psychology realizes to a greater degree the importance of establishing such facts with accuracy.

In this group, also, fall those findings which tend to establish a new technique of procedure in researches of the type here undertaken. In view of these facts we draw out of the body of this report detailed conclusions, organized so as to bring out their real significance.

1. Reported preference within a narrow field can be measured with a remarkable degree of accuracy by employing definite reporting methods in narrowed-down preference situations.

2. Reported preference in the particular connection here studied correlates with performance to the degree of .56.

3. This correlation compares somewhat favorably with that existing between performance and Otis IQ's, which we found to be .63.

4. This correlation is somewhat larger than that between performance as here measured and reading ability as measured by Gates' test.

5. This relation is not large enough to predict performance from reported preference, as is indicated by the P. I. of the r. In this particular field, a glib prediction as to the probable success of a pupil, knowing his reported preference, is entirely unwarranted.

6. Pupils of high ability tend to report no distinct preference as to type of problem and tend to perform more nearly equally well on all types. This, consequently, evidences a lower degree of relationship between preference and performance. This does not mean that pupils of high ability have no preferences, but that these preferences are determined by a variety of causes and not merely by the presence or absence of computation difficulties. These

preferences, therefore, tend to have little consistency as to types preferred.

7. Difference is evident between the grades in their reported preference and performance. As you go from the higher grades to the lower, you can observe a tendency to prefer and to perform to a less extent the problems requiring analyses. The lower grades more distinctly prefer the computation type, while the pupils of higher grades prefer it to a less degree.

8. Pupils of lower ability report preferences more consistently for problems involving little or no complex situations. Also they perform relatively better on these than on those requiring more involved analyses. (The analyses here referred to are analyses of situations described in problems.) Lower ability pupils, consequently, prefer problems of pure computation with definite directions stated.

9. Boys are superior to girls in performance in arithmetic problem solving as measured in this study.

10. As measured here girls show a high reported preference and performance with respect to problems of the puzzle and pure computation type, whereas the boys reveal a greater degree of performance and preference with respect to the adult, child, and science types.

11. The degree of consistency in choosing a particular type appears to be due to a considerable extent to the rather uniform choice of the computation type by pupils of relatively low ability. Should this type have been omitted, it is probable that the remaining four types would have been ranked in the same way with respect to performance and preference, and that there would have been a lower degree of consistency in preference.

12. Of problems involving descriptive situations, the child type is most preferred and most successfully performed by all groups of junior high school pupils. The adult type ranks second, while the science type ranks lowest in both preference and performance. We may infer that this ranking is probably due to the pupils' relative experience with the objects and situations involved in the three types of problems, respectively.

13. The inference may be drawn from the above findings and conclusions that preferences for, or so-called interests in, certain problems are probably occasioned to a certain extent by the pupils' experiences which have been built up, and that they are influenced by whatever growth and changes occur in the experiences. The evidence as found in this study seems to justify the inference that with respect to problem solving by pupils of the junior high school grades, preferences are not permanent and inherent qualities, but are somewhat dependent upon estimated ability resulting from experience with the respective situations involved.

PRACTICAL CONCLUSIONS

This study also has significant practical conclusions. We know that the content of junior high school mathematics is yet largely undetermined. It

is, therefore, of practical significance to know what types of problems these students prefer, on which they perform best, and what the relation of preference and performance is. Some of the more useful findings are here summarized:

1. The general intelligence of pupils should serve as a basis for the selection of arithmetic problems for the junior high school. Both the relative degrees of performance and of reported preference on the various types of problems vary as between levels of general intelligence.

2. The relation between reported preference and performance is significant enough to demand some attention when we select problems for text books.

3. For lower intelligence levels, exercise of the pure computation, and drill in fundamentals seem to be most highly preferred.

4. Boys and girls differ in both performance and preference with respect to types of problems. There seems to be justification, then, in taking this into account when presenting problems to the two sexes. Reasoning problems of less difficulty in analyses should probably be provided for the girls, while the boys might profitably be given those of greater complexity.

5. The pupils of the lower intelligence levels indicate little preference for and ability in problems of the science type and the puzzle type. Such problems should be constructed on a lower ability level when presented to pupils of this type.

6. Pupils of high ability in problem solving and in intelligence show a comparatively high degree of performance and preference with respect to problems of the puzzle and the science type.

7. The science type ranked lowest in performance and preference by the group as a whole. The relatively greater degree of preference and performance with respect to this type shown by pupils of higher ability, however, suggests that more problems taken from the sciences might profitably be included in an arithmetic course of study. If we wish to pitch problem solving more nearly on the ability level of the upper decile, for example, we should select many problems of the science type.

8. Problems dealing with child activities seem to be highly preferred and performed by all pupils of the junior high school grades. We are therefore justified in believing that if arithmetic problems, through their solutions, offer to children means of meeting their real situations, they become very vital factors in the education of the child. These problems must be built up on childhood situations and should not be merely about childhood and child life.

9. Building up interest in problem solving is best accomplished by having children work problems which are not too difficult and which represent genuine childhood situations. This is especially true in the case of pupils who are just at or below the median ability of the class.

Further Needed Studies

In view of the findings in this investigation, the following studies are suggested:

1. The influence of training and experience upon preferences should be studied.
2. The relation of reported preference to performance in other fields than arithmetic should be established.
3. By comparison between relationships found in various fields, the determination of some of the principles which underlie the formation of preferences should be accomplished.
4. A more careful study of the reasons for the variation in performance with respect to different descriptive problems when the computation difficulties are held constant, is highly desirable.
5. The social factors that influence preference should be determined.
6. The particular and genuine childhood arithmetic situations upon which problems for pupils of the junior high school level can be built should be discovered.

BIBLIOGRAPHY

Bailor, Edwin M., "Content and Form in Tests of Intelligence," *Teachers College, Columbia University, Contributions to Education*, Number 162, 1924.

Bridges, J. W., and Dollinger, V. M., "The Correlations Between Interests and Ability in College Courses," *Psychological Review*, 1920, Volume 27, pp. 308-314.

Columba, Sister M., "A Study of Interests and their Relation to Other Factors of Achievement," *Catholic University of America, Educational Research Bulletin*, 1925, Number 1, pp. 1-55.

Fryer, Douglas, "Predicting Ability from Interests," *Journal of Applied Psychology*, 1927, Volume 11, pp. 212-225.

Garrett, Henry E., *Statistics in Psychology and Education*, First Edition, Longmans, Green and Company, 1926.

Hubbard, Ruth M., "A Measurement of Mechanical Interests", *Journal of Genetic Psychology*, June 1928, Volume 35, pp. 229-254.

Thorndike, Edward L., "Early Interests, Their Permanence and Relation to Abilities," *School and Society*, Volume 5, pp. 178-179.

Thorndike, Edward L., *The Psychology of Arithmetic*, The Macmillan Company, 1922, p. 22.

Uhrbrock, Richard S., "Interests as an Indication of Ability," *Journal of Applied Psychology*, 1926, Volume 10, pp. 487-501.

VITA

Herbert Lloyd Bowman

Born February 27, 1888, in Perry County near Perryville, Missouri.

Education: Rural Schools, Perry County, Missouri, 1894-1907; Southeast Missouri State Teachers College, 1908-1917; University of Wisconsin, summer terms, 1923 and 1924; University of Missouri, 1926-1929 and summer of 1929. M. A. degree, 1927, Ph.D., 1929.

Teaching experience: Four years rural teacher, Perry County, Missouri, 1908-1912; eleven years principal and superintendent in public schools of Missouri during the years 1912-1926; instructor in education, Southeast Missouri State Teachers College, summer 1927; instructor in education, University of Missouri, Rolla summer session, 1928.